

# Decorrelation by recurrent inhibition in heterogeneous neural circuits

Alberto Bernacchia and Xiao-Jing Wang\*

November 17, 2011

## Abstract

The activity of cortical neurons is correlated, and this correlation affects how the brain processes information. What are the neural circuit mechanisms underlying correlations? Shared synaptic input is necessary to synchronize neurons, but it may not be sufficient if neural circuits are heterogeneous. We study the neural circuit mechanisms of correlations by analyzing a network model characterized by strong and heterogeneous interactions. We consider strong excitatory and inhibitory connections: the excitatory input drives the fluctuations of neural activity, which are counterbalanced by the inhibitory feedback. While the strong excitatory input tends to correlate neurons, the inhibitory feedback strongly reduces the correlations. Crucially, the heterogeneity of synaptic connections is necessary for the inhibition of correlations. Under these conditions, we found that correlations are positive and of magnitude  $K^{-\frac{1}{2}}$ , where  $K$  is the number of connections to a neuron. The result agrees with anatomical and physiological measurements in the cortex: a cortical neuron receives about  $K \simeq 10^4$  connections, and the measured correlations are about  $10^{-2}$ .

## Introduction

Simultaneous measurements of the activity of multiple neurons have shown significant correlations, and this observation has stimulated the debate on whether and how correlations contribute to neural computation. In principle, correlations allow robust signal processing, because redundancies across neurons can be exploited to separate the signal from the noise [1, 31]. Experimental studies of the cerebral cortex suggest that correlations improve decoding of stimuli [18], but it remains unclear whether a parsimonious decoder should rely on correlations [2]. A challenge to this hypothesis is the observation that correlations are reduced when

---

\*Department of Neurobiology, Yale University

animal subjects are actively engaged in discrimination [9, 10], and even when they simply start moving [32]. In addition, neurons with similar responses to stimuli show higher correlations [44, 26, 27, 4, 12, 2, 36, 23, 39, 22, 15, 24], implying that discrimination of stimuli is worsened by correlations [1, 31, 40, 42, 3, 19]. Another caveat is that the neural code is largely unknown, and if the "noise" measured in physiological studies encodes some signal then any correlation would decrease the available information [29].

Besides the possible function of correlations in signal and information processing, their physiological causes remain unclear. It has been shown that the correlation between nearby neurons is driven by their shared synaptic input [25, 32]. However, a quantitative understanding of the circuit mechanisms regulating correlations between cortical cells is still missing, and the goal of this study is to determine the dependence of correlations on different properties of the neural circuitry. The measured correlation between neurons depends on different factors and varies across studies [11]: it increases with the proximity of neurons [27, 12, 39, 15, 24], their activity [13] and the timescale on which the activity is integrated [4, 34, 12, 2, 23, 39, 22, 28]. Here we do not consider the effects of distance and firing rate, and we study the correlation of the instantaneous activity, calculated on the timescale of a single action potential (about 1ms). Fig.1 shows the correlation measured in eight different studies as a function of timescale; When measured at short timescales, the correlation is positive and of the order of  $10^{-2}$ .

Previous modeling studies of neural circuits have found that the mean correlation between neurons is small, of the order of  $N^{-1}$ , where  $N$  is the number of neurons in the network. Small correlations have been observed, not surprisingly, in networks characterized by weak connection strengths [17, 5]. More surprisingly, the same result has been obtained in the case of strong connections, the high-conductance state [14, 41], provided that the network includes a strong inhibitory feedback [35, 21]. Here we provide an analytical study of correlations, and we confirm both the observed small correlation, and the crucial effect of the inhibitory feedback in reducing it; However, we find that correlations are of magnitude  $K^{-\frac{1}{2}}$ , where  $K$  is number of connections received by a neuron, instead of the  $N^{-1}$  observed in previous studies. We argue that our result is consistent with empirical observations, since a cortical neuron receives about  $K \simeq 10^4$  connections and the correlation is  $\sim 10^{-2}$ . In addition, we show that this result crucially depends on the heterogeneity of connection strengths.

## Methods

The model consists of a neural circuit of  $N$  neurons, receiving input from  $N_{ext}$  external neurons. The dynamics of neuronal activity is described by a linear firing rate model: each neuron integrates the signal from other neurons, weighted

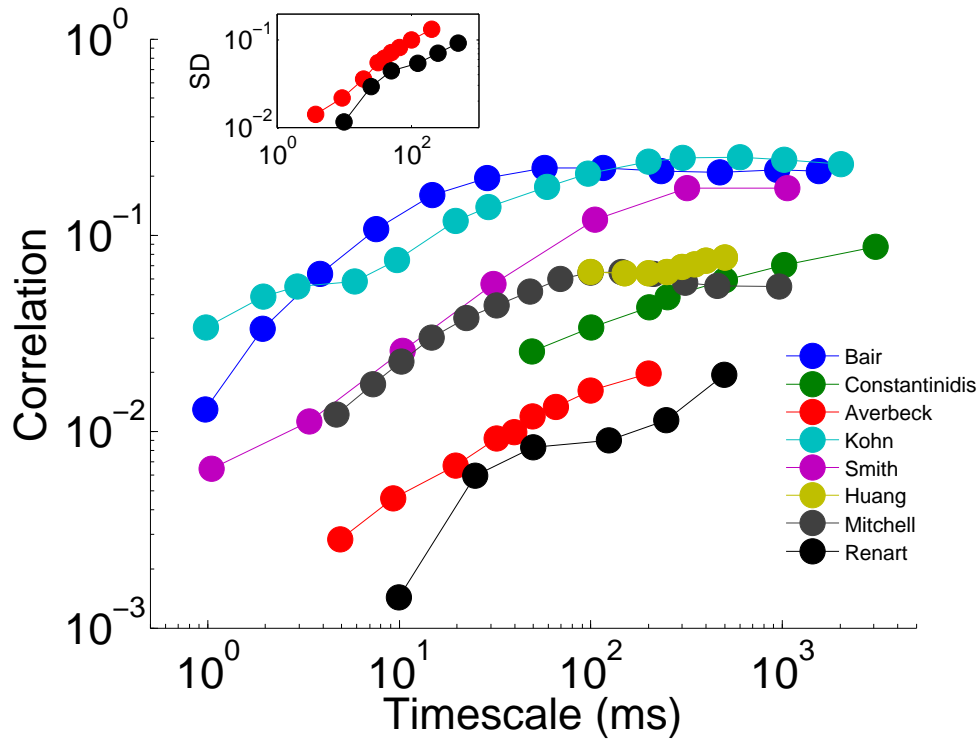


Figure 1: Mean correlation across neuron pairs plotted vs the timescale on which correlations are measured. Re-plotted from eight experimental studies of the cortex (legend). Two studies provided not only the mean correlation but also the Standard Deviation (SD, inset).

by the synaptic connection strength. The dynamics of the model is described by the equation

$$\tau \frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^N G_{ij} x_j(t) + \sum_{j=1}^{N_{ext}} G_{ij}^{ext} x_j^{ext}(t) \quad (1)$$

where  $x_i$  is the activity of neuron  $i$  in the local circuit, and  $G_{ij}$  is the strength of the synaptic connection from neuron  $j$  to neuron  $i$ . The external (feed-forward) input to the circuit is provided by the activities  $x_j^{ext}$ , and the synaptic connection from the  $j$ -th external neuron to the  $i$ -th local neuron is given by the strength  $G_{ij}^{ext}$ . All neuronal activities evolve in time, while the connectivity matrices  $G$  and  $G_{ext}$  are fixed.

We define the average number of local connections received by a neuron as  $K$ , and the external connections as  $K_{ext}$ . We assume that the connectivity matrices are random and we consider two scenarios (represented schematically in Fig.2a,b):

- 1) The network is fully connected ( $K = N$ ,  $K_{ext} = N_{ext}$ ) with random connection strengths (All-to-All, Fig.2a), characterized by a Gaussian distribution.
- 2) The network is sparse, only a fraction of connections exists ( $k = K/N$ ,  $k_{ext} = K_{ext}/N_{ext}$ ), selected at random but of constant strength (Sparse, Fig.2b), characterized by a Bernoullian distribution.

The randomness of connections makes the network akin to a "disordered" system, which is characterized by a random but fixed substrate. We adopt a single notation for either case, All-to-All or Sparse network, by defining the mean and variance of matrix elements and their scaling with  $K$ ,  $N$ :

$$\langle G_{ij} \rangle = -kg/\sqrt{K} \quad \langle \Delta G_{ij}^2 \rangle = \lambda^2/N \quad (2)$$

$$\langle G_{ij}^{ext} \rangle = k_{ext}g_{ext}/\sqrt{K_{ext}} \quad \langle \Delta G_{ij}^{ext2} \rangle = \lambda_{ext}^2/N_{ext}. \quad (3)$$

(angular brackets denote average over the matrix distribution,  $\Delta$  indicates variation around the mean). In the All-to-All network,  $k = 1$  and  $K = N$ , and the parameters of the matrix distribution are  $g$  and  $\lambda$ . In the Sparse network, connection strengths have only two possible values, either zero or  $-g/\sqrt{K}$ , the two parameters are  $g$  and  $k$ , but for convenience of notation we also use the parameter  $\lambda^2 = g^2(1 - k)$ . Similar equations and parameters describe the external connections in either case. The mean connection is negative for  $G$  (inhibitory) and positive for  $G_{ext}$  (excitatory), since  $g$  and  $g_{ext}$  are positive. Theoretical analysis also considers the case of local excitatory and inhibitory populations (in which  $\langle G \rangle$  is a generic rank-1 matrix, see Appendix 2).

We assume that the external activity  $\mathbf{x}_{ext}(t)$  is a stochastic process uncorrelated in both space and time, i.e. a white noise characterized by mean  $\overline{x_i^{ext}(t)} =$

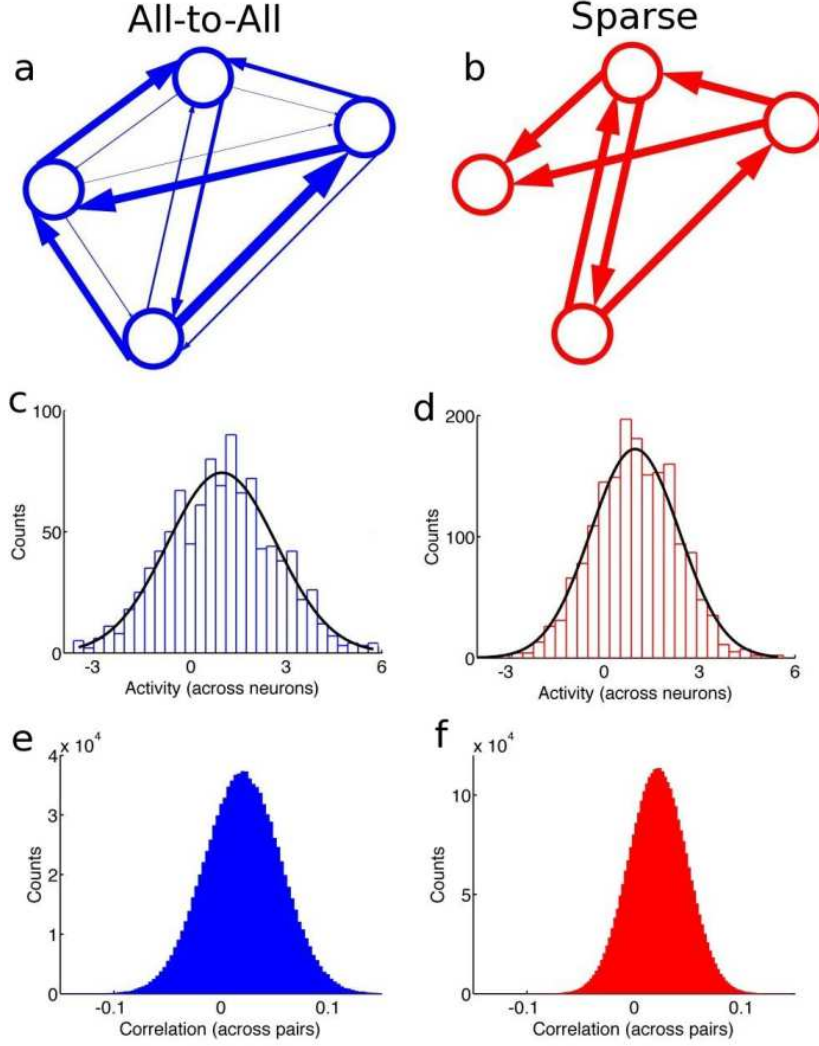


Figure 2: Scheme of the network and distribution of activity and correlations. We study two network architectures: All-to-All connectivity with random strengths (a), and Sparse random connections of fixed strength (b). Connection strength is illustrated by the thickness of edges. The distribution of activity across neurons (c,d) and the distribution of correlations across neuron pairs (e,f) are Gaussian for both types of networks ((c,e) for the All-to-All network, (d,f) for the Sparse network). The parameters used are:  $g = g_{ext} = 1$ ,  $\bar{x}_{ext} = \overline{\Delta x_{ext}^2} = 1$ ,  $K_{ext} = K$ . For the All-to-All network:  $K = 1000$ ,  $\lambda_{ext} = 1$ ,  $\lambda = 1/\sqrt{2}$ . For the sparse network:  $K = 880$  and  $k_{ext} = k = 1/2$ , which correspond to  $\lambda = \lambda_{ext} = 1/\sqrt{2}$ .

$\overline{x}_{ext}$  and covariance  $\overline{\Delta x_i^{ext}(t)\Delta x_j^{ext}(t')} = \overline{\Delta x_{ext}^2}\delta_{ij}\delta(t-t')$  (overline denotes the average over different realizations of the noise, and  $\delta$  denotes either the discrete Kronecker or continuous Dirac function). Therefore, Eq.(1) corresponds to a Ornstein-Uhlenbeck stochastic process [16].

## Results

We study neural activity and correlations among neurons in a heterogeneous neural circuit model. Local recurrent connections are dominated by inhibition, while external feed-forward projections are excitatory. We consider two types of circuits, All-to-All connectivity with random strengths (Fig.2a) and Sparse random connections of fixed strengths (Fig.2b). Results are displayed in a single notation for either case (see Methods). Fig.2c,d shows the distribution of activity across neurons, and Fig.2e,f shows the distribution of correlations across neuron pairs. The purpose of this work is to describe how the mean and variance of those distributions depend on the parameters of the neural circuit.

Since the model is linear, the neural activity  $x$  can assume negative values, while the activity of real neurons is positive (firing rate); Those values are interpreted as deviations from a steady state, attained by the network in absence of the external input, around which the neural dynamics is approximately linear. If we denote the steady activity as  $f_0$ , the firing rate is equal to  $f = f_0 + cx$ , where we assume that  $f$  is positive and  $c$  is a positive constant. As long as the linear approximation is valid, the correlations observed in the model are insensitive to the nature of the steady state (i.e. to the values of  $f_0$  and  $c$ ).

Due to the linearity of the model, all quantities of interests can be simply calculated; The novel contribution of this work is averaging those quantities over the randomness of the connectivity matrix. Because connections are heterogeneous, different neurons have a different activity, and we compute the sample mean across neurons in order to obtain the spatial average. If the number of neurons  $N$  is large, this is independent on the specific realization of the connectivity, therefore we perform its average over the distribution of connections, and we obtain (see Eq.(13) in Appendix 1; Angular brackets denote averaging over the random connectivity, overline denotes temporal average)

$$\langle \overline{x} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \overline{x}_i \right\rangle = \frac{g_{ext}\sqrt{K_{ext}}}{1 + g\sqrt{K}} \overline{x}_{ext} \quad (4)$$

The numerator of this expression is equal to the mean excitatory input received by a neuron,  $g_{ext}\sqrt{K_{ext}} \overline{x}_{ext}$ , while the denominator expresses the recurrent inhibition, whose total post-synaptic strength is  $g\sqrt{K}$ . Therefore, the strong recurrent inhibition counterbalances the large excitatory input and determines a relatively low activity, regardless of the network size. Note that the number of local and external connections,  $K$  and  $K_{ext}$ , are both large, but they tend to balance in

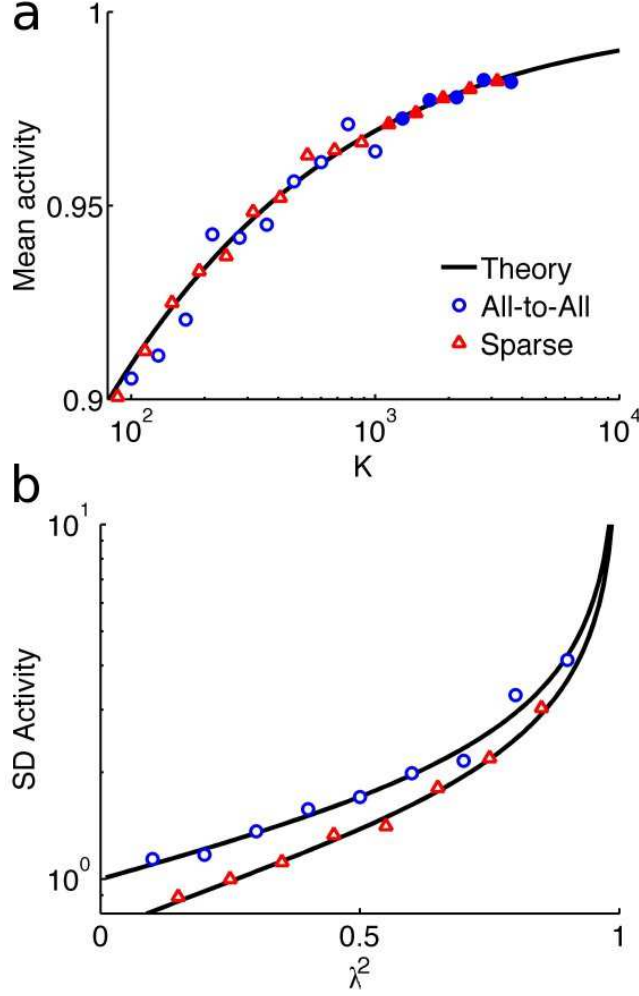


Figure 3: Mean (a) and Standard Deviation (b) of activity across neurons. The mean is plotted against the number of connections  $K$ , and the SD is plotted against the heterogeneity  $\lambda$ . Analytical results (lines) are obtained from Eq.(4) in panel a and Eq.(5) in panel b. Simulation results are shown for the All-to-All (blue circles) and Sparse (red triangles) network. Empty symbols show simulations of the neural dynamics, Eq.(1); Filled symbols show numerical evaluation of Eq.(11). The parameters used are, in both panels,  $g = g_{ext} = 1$ ,  $\bar{x}_{ext} = \overline{\Delta x_{ext}^2} = 1$ . In panel b,  $K_{ext} = K = 500$ . For the All-to-All network:  $\lambda_{ext} = 1$ , and  $\lambda = 1/\sqrt{2}$  in panel a. For the Sparse network:  $k_{ext} = k = 1/2$  in panel a (which correspond to  $\lambda = \lambda_{ext} = 1/\sqrt{2}$ ), while  $k_{ext} = k$  is varied in panel b.

the expression above. Fig.3a shows an example of mean activity as a function of the number of connections; The mean activity is rather insensitive to the number of connections, which are taken equal to the external ones in each simulation ( $K = K_{ext}$ ). The analytical result, Eq.(4), agrees with numerical simulations of the network dynamics, in both the All-to-All and the Sparse networks.

Different neurons have different connections and therefore different activity, and the extent to which the activity varies from neuron to neuron is determined by the spatial variance. We calculate this quantity by taking the sample variance across neurons and averaging over the random connectivity, and we obtain (see Eq.(16) in Appendix 1)

$$\langle \Delta \bar{x}^2 \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \Delta \bar{x}_i^2 \right\rangle = \frac{1}{1 - \lambda^2} \left[ \langle \bar{x} \rangle^2 \lambda^2 + \bar{x}_{ext}^2 \lambda_{ext}^2 \right] \quad (5)$$

The spatial variance of neural activity increases with the network heterogeneity, expressed by  $\lambda$  and  $\lambda_{ext}$  for, respectively, the recurrent and external connections. Increasing the heterogeneity of connections increases the differences in the total input between neurons and therefore in their activities. The spatial variance is also proportional to the mean activity, local  $\langle \bar{x} \rangle$  and external  $\bar{x}_{ext}$ . Furthermore, increasing the heterogeneity of recurrent connections leads to a divergence of the spatial variance, when  $\lambda^2$  approaches one. In this case the state  $x = 0$ , which we assumed to correspond to a steady state of firing rates, destabilizes, and the linear approximation fails (see Appendix 1). Fig.3b shows an example of the spatial variance as a function of the variability of the recurrent connections. The analytical result, Eq.(5), agrees with numerical simulations of the network dynamics, in both the All-to-All and the Sparse network.

After looking at the mean and spatial variability of neural activity, we turn to the main theme of our work, the analysis of temporal variability and correlations. The activity of each neuron fluctuates in time, due to the fluctuating input, and those temporal fluctuations may be correlated since different neurons receive shared input. We study temporal variability and correlated fluctuations by calculating the covariance matrix, in particular the instantaneous covariance, at zero time lag. First, we look at the on-diagonal elements of this matrix, which are the temporal variances of different neurons. To determine the average temporal variance, we take the sample mean across neurons and we average over the random connectivity, obtaining (see Eq.(19) in Appendix 1)

$$\langle \overline{\Delta x^2} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \overline{\Delta x_i^2} \right\rangle = \frac{\overline{\Delta x_{ext}^2}}{2} \left[ \frac{k_{ext} g_{ext}^2}{1 + g\sqrt{K}} + \frac{\lambda_{ext}^2}{\sqrt{1 - \lambda^2}} \right] \quad (6)$$

The temporal variance of neural activity is the sum of two pieces: the first term decreases with the number of connections as  $K^{-1/2}$ , while the second term remains finite. The first term indicates that recurrent inhibition ( $g$ ) reduces temporal fluctuations. In fact, the inhibitory feedback not only reduces the mean activity



(Eq.(4)), but also cuts down fluctuations by quickly counterbalancing the external excitatory input. This can be verified by calculating the instantaneous covariance between the external excitatory and the local inhibitory input, which is found large and negative, equal to  $-k_{ext}g_{ext}^2g\sqrt{K}$ . Furthermore, as in the case of spatial fluctuations, temporal fluctuations increase with the heterogeneity of connections, recurrent ( $\lambda$ ) and external ( $\lambda_{ext}$ ). Temporal fluctuations also diverge when the network approaches the instability point, when the linear approximation fails ( $\lambda \rightarrow 1$ ).

How much of the total variance, expressed in Eq.(6), is independent rather than shared between neurons? To answer this question we calculate the average covariance, by looking at the off-diagonal elements of the covariance matrix, the pairwise covariances. We take the sample mean across neuron pairs and average over the random connectivity to obtain the average covariance (see Eq.(20) in Appendix 1)

$$\langle \overline{\Delta x \Delta x} \rangle = \left\langle \frac{1}{N(N-1)} \sum_{i \neq j}^{1,N} \overline{\Delta x_i \Delta x_j} \right\rangle = \frac{\overline{\Delta x_{ext}^2}}{2} \frac{k_{ext}g_{ext}^2}{1 + g\sqrt{K}} \quad (7)$$

Notably, this is equal to the first term in the total variance, Eq.(6), implying that the two terms in the total variance express, respectively, the correlated and uncorrelated fluctuations. Therefore, while the uncorrelated variance remains finite for large  $K$ , the correlated variance vanishes. The activities of neuron pairs tend to covary, due to their shared external input, but the recurrent inhibition makes the covariance small, of order  $K^{-1/2}$ . In the Sparse network, the covariance vanishes if the probability of external connections is small ( $k_{ext} \rightarrow 0$ ), since the shared external input between neurons tends to zero in that case. In the All-to-All network, the mean covariance vanishes if the mean input connection is zero ( $g_{ext} = 0$ ); In that case, even if neurons receive a shared external input, neuron pairs may weight different inputs with the same or opposite signs, leading to respectively positive or negative covariance. Therefore, while the mean covariance across neuron pairs is zero, the covariance of single pairs may be positive or negative.

The mean correlation is obtained by dividing the covariance, Eq.(7), by the variance, Eq.(6), i.e. (we assume that variance and covariance are independent):

$$\langle R \rangle = \frac{\langle \overline{\Delta x \Delta x} \rangle}{\langle \overline{\Delta x^2} \rangle} = \frac{1}{1 + \frac{\lambda_{ext}^2}{\sqrt{1-\lambda^2}} \frac{1+g\sqrt{K}}{k_{ext}g_{ext}^2}} \quad (8)$$

This expression is positive and never exceeds one. It indicates that the mean correlation is small, of order  $K^{-1/2}$ , despite the strong and shared excitatory input between neurons. However, this result holds only in presence of the local recurrent inhibition ( $g > 0$ ), and provided that external connections are heterogeneous ( $\lambda_{ext}^2 \neq 0$ ). Heterogeneity of local connections ( $\lambda$ ) also contributes in decreasing

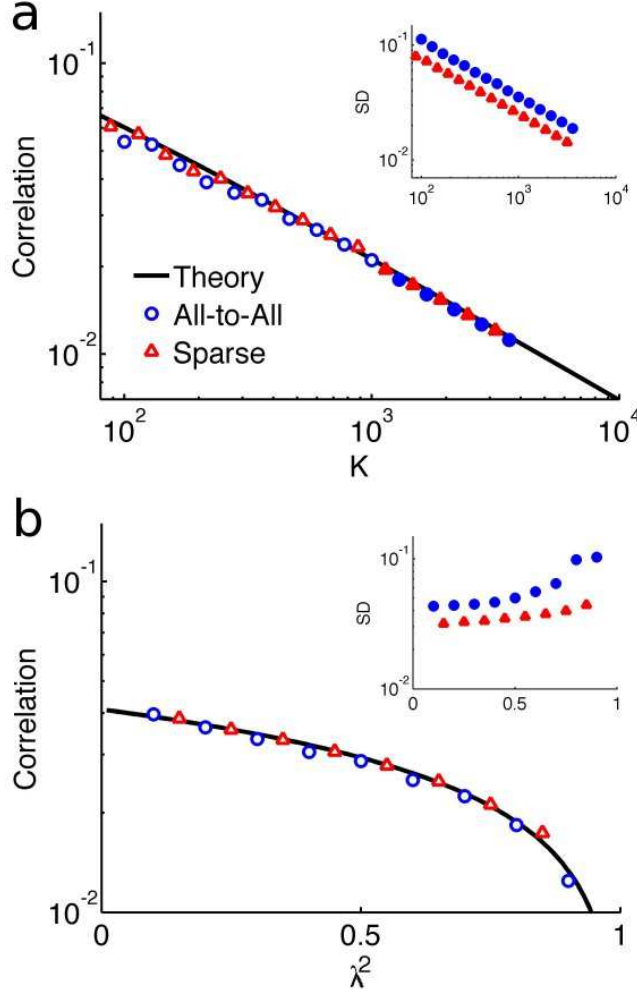


Figure 4: Mean correlation in the activity between neuron pairs, Standard Deviation is shown in the insets. The correlation is plotted against the number of connections  $K$  (a) and the heterogeneity  $\lambda$  (b) of the network. Analytical results (lines) are obtained from Eq.(8). Simulation results are shown for the All-to-All (blue circles) and Sparse (red triangles) network. Empty symbols show simulations of the neural dynamics, Eq.(1); Filled symbols show numerical evaluation of Eq.(17). The parameters used are, in both panels,  $g = g_{ext} = 1$ ,  $\bar{x}_{ext} = \overline{\Delta x_{ext}^2} = 1$ . In panel b,  $K_{ext} = K = 500$ . For the All-to-All network:  $\lambda_{ext} = 1$ , and  $\lambda = 1/\sqrt{2}$  in panel a. For the Sparse network:  $k_{ext} = k = 1/2$  in panel a, which correspond to  $\lambda = \lambda_{ext} = 1/\sqrt{2}$ , while  $k_{ext} = k$  is varied in panel b.

the correlation.

Therefore, the inhibitory feedback and the random connectivity are responsible for the small correlation. If the inhibitory feedback is removed,  $g = 0$ , the correlation becomes large. If the network heterogeneity is removed,  $\lambda = \lambda_{ext} = 0$ , the correlation is equal to one, because the network is homogeneous and all neurons get the same input (but see Appendix 1 for a caveat on this case, Eq.(22)). Fig.4 shows an example of the mean correlation as a function of the number of connections and the heterogeneity of the network. The analytical result, Eq.(8), agrees with numerical simulations of the network dynamics, in both the All-to-All and the Sparse network. Insets in Fig.4 show the Standard Deviation of correlations, which appear to decrease with the number of connections as  $K^{-1/2}$ , and to increase with the heterogeneity of the network.

For the Sparse network, it is interesting to note that all the above results depend on the number of neurons  $N$  only through the parameter  $\lambda^2 = g^2(1 - K/N)$ . Therefore, while the above quantities may depend crucially the number of connections  $K$ , they depend only weakly on the number of neurons  $N$ . In particular, the order of magnitude of correlations  $K^{-1/2}$  holds regardless of the number of neurons, which may be taken even infinite for any fixed value of  $K$  (provided that  $g < 1$  and  $k_{ext}$  is non-zero). While in other studies the results are often described in terms on the number of neurons  $N$ , here the determinant variable is the number of connections  $K$ .

## Discussion

We found that inhibitory feedback and heterogeneous connections have important effects on the dynamics of the activity in a neural circuit. The strong excitatory input, shared between neurons, tends to drive the network to a highly active and correlated state. The inhibitory feedback is responsible for balancing the network activity, and also for reducing temporal fluctuations, in particular the correlated fluctuations across neurons. The heterogeneity of couplings also plays a crucial role in reducing correlations, since homogeneous connections would determine homogeneous and therefore highly correlated activity. As a consequence, the observed mean correlation is positive and of small magnitude. The fact that mean correlation is positive is obvious, since neurons in a large population cannot be anti-correlated on average<sup>1</sup>. What is not obvious is that the mean correlation is of small magnitude.

In presence of heterogeneous connections and inhibitory feedback, the mean

---

<sup>1</sup>The mean correlation must be larger than  $-\frac{1}{N-1}$ , therefore it must be positive in infinitely large populations. Proof: Any covariance matrix  $Q$  is positive definite, therefore  $h^\dagger Q h > 0$  for any vector  $h$ . If we choose  $h_i = 1/\sqrt{Q_{ii}}$  and define the correlation matrix  $R_{ij} = Q_{ij}/\sqrt{Q_{ii}Q_{jj}}$  we have that  $\sum_{ij} R_{ij} = N + N(N-1)\langle R \rangle > 0$ . Therefore the mean correlation must be  $\langle R \rangle > -\frac{1}{N-1}$

correlation decreases with the number of connections as  $K^{-\frac{1}{2}}$ . This order of magnitude agrees with experimental observations: a cortical neuron receives about  $K \sim 10^4$  connections [7] and the measured mean correlation is of order  $10^{-2}$ . Note that most experimental studies report a higher value of the mean correlation in the mammalian cortex, around  $10^{-1}$  [11]. However, different studies measure correlations at different timescales, by integrating activity (or correlations) on different time windows. Larger temporal windows imply larger covariance, because neurons covary on multiple timescales, and the increase in covariance is not accompanied by a comparable increase in the variance, due to after-hyperpolarization effects [4]. Therefore, a significant increase in the correlation, of about one order of magnitude, is observed when increasing the time window from about  $1ms$  to  $1s$  (see Fig.1). Here we consider only instantaneous correlations: when measured at a short timescale ( $\sim 1ms$ ), the correlation is about  $10^{-2}$ . Besides the cerebral cortex, our model may be useful to investigate correlations in other brain areas, especially those dominated by inhibition, such as the striatum and globus pallidus. Interestingly, large correlations between pallidal neurons have been recently proposed as the origin of Parkinsonism [33, 43].

Other modeling studies have addressed the issue of correlations in neural circuits. In the studies [35, 21], the mean correlation was found to decrease with the number of neurons as  $N^{-1}$  instead of the order  $K^{-\frac{1}{2}}$  that we found here. One difference between those models and ours is that we study a firing rate model instead of a spiking model, but that may not be a crucial difference. In those studies, the small correlation has been explained by an exact and instantaneous tracking of the excitatory input by recurrent inhibition. In particular, [35] points out that instantaneous tracking is the key for the  $N^{-1}$  scaling of correlations, and if tracking is delayed then correlations would be of the order  $N^{-\frac{1}{2}}$  instead. In our model, although recurrent inhibition is largely anti-correlated with the excitatory input, tracking is not instantaneous, since the excitatory input and the recurrent inhibition fluctuate on different timescales; Excitation is fast, while inhibition is slowed down by the recurrent dynamics. This difference may be responsible for the higher order of correlation observed in our study.

While neural dynamics occurs on a variety of timescales in our model<sup>2</sup>, as well as in real neurons [6], it may not be fast enough for instantaneous tracking. It would be instructive to slow down the fluctuations of the excitatory input in our model, for example by injecting colored instead of white noise in the network. In biological networks however, it is not known whether inhibition can track excitation exactly and instantaneously, as described in [35, 21]. Strong anti-correlations between excitatory and inhibitory inputs have been observed experimentally [30, 8, 37], but it remains unclear to what extent they may decorrelate neural activity.

---

<sup>2</sup>The network is endowed with one very fast timescale,  $g^{-1}K^{-1/2}$ , while the remaining timescales are in between  $(1 + \lambda)^{-1}$  and  $(1 - \lambda)^{-1}$ , in time units of  $\tau$ .

The difference in the order of magnitudes observed in our work compared to previous modeling studies may seem subtle, but this difference may be substantial from a functional standpoint. The magnitude of correlations found in our study is in between the "asynchronous" state [17, 35], in which the mean correlation is  $\sim N^{-1}$ , and the "synchronous" state, in which correlations are of order one. Those two models are opposite to one another, not only from a phenomenological point of view, but also in regard to the interpretation of the neural code and the hypotheses on the mechanisms underlying brain function. The former predicts that information is transmitted in the brain through firing rates, while the latter suggests that is transmitted through synchronization of neural populations. There could be some truth in both views, but a satisfactory model bridging the two is not yet available. Further studies of correlations may provide the necessary understanding to fill the gap.

## Acknowledgements

This study was supported by the US National Institutes of Health grant R01 MH062349 and the Swartz Foundation.

## Appendix 1: Statistics of random networks

In this section, we calculate the averages of neural activity and correlations with respect to both temporal fluctuations (noise) and the spatial variability of the connection strengths (disorder). Due to the linearity of the model, all quantities of interests can be simply calculated; The novel contribution of this work is averaging those quantities over the randomness of the connectivity matrix. The equation of dynamics (1) can be expressed in matrix form (the time constant of temporal evolution  $\tau$  is set to 1):

$$\frac{d\mathbf{x}(t)}{dt} = (G - I)\mathbf{x}(t) + G_{ext}\mathbf{x}_{ext}(t) \quad (9)$$

where  $\mathbf{x}$  is the vector of local neural activities,  $\mathbf{x}_{ext}$  is the vector of external neural activities, the matrices  $G$  (of size  $N \times N$ ) and  $G_{ext}$  (size  $N \times N_{ext}$ ) express respectively the recurrent connections and the feed-forward projections, and  $I$  is the identity matrix. The equation of dynamics is linear and, given the interaction matrices  $G$ ,  $G_{ext}$  and the input signal  $\mathbf{x}_{ext}$ , the neural activity can be expressed as a sum over the external input weighted by an exponential temporal decay

$$\mathbf{x}(t) = \int_{-\infty}^t dt' e^{(G-I)(t-t')} G_{ext}\mathbf{x}_{ext}(t') = \int_0^{+\infty} dt' e^{(G-I)t'} G_{ext}\mathbf{x}_{ext}(t-t') \quad (10)$$

We assumed that initial conditions have decayed and that the inequality  $\lambda < 1$  holds, to prevent network activity from growing in time without bounds<sup>3</sup>.

We start by calculating the mean neural activity. We perform the temporal average of the above expression, therefore we substitute the external activity  $\mathbf{x}_{ext}(t)$  with its average  $\bar{x}_{ext}$ , and we perform the integral, obtaining (temporal average is denoted by overline)

$$\bar{\mathbf{x}} = \bar{x}_{ext} (I - G)^{-1} G_{ext} \mathbf{1} \quad (11)$$

where the vector  $\mathbf{1}$  has all  $N_{ext}$  components equal to one. Because the matrices of connection strengths are heterogeneous,  $G$  and  $G_{ext}$ , different neurons have a different mean activity. In order to calculate the spatially averaged activity, we compute the sample mean across neurons. For large  $N$ , this is independent on the specific realization of the spatial disorder, therefore we perform its average over the distribution of connectivity strengths, namely

$$\langle \bar{x} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \bar{x}_i \right\rangle = \left\langle \frac{\bar{x}_{ext}}{N} \mathbf{1}^\dagger (I - G)^{-1} G_{ext} \mathbf{1} \right\rangle \quad (12)$$

The average (angular brackets) is across all possible realizations of the random matrices  $G$  and  $G_{ext}$ . We denote by  $\dagger$  the transpose operation. Note that, in the expression above, the row vector  $\mathbf{1}^\dagger$  has  $N$  components, while the column vector  $\mathbf{1}$  has  $N_{ext}$ . In the following we will use the same notation regardless of the dimension of  $\mathbf{1}$ , since that can be determined by the dimension of the multiplied matrix. Since  $G$  and  $G_{ext}$  are independent, we can substitute  $G_{ext}$  with its mean,  $\langle G_{ext} \rangle = \frac{g_{ext} \sqrt{K_{ext}}}{N_{ext}} \mathbf{1} \mathbf{1}^\dagger$ . Furthermore, we show in Appendix 2, Eq.(36), that  $\langle \mathbf{1}^\dagger (I - G)^{-1} \mathbf{1} \rangle = N(1 + g\sqrt{K})^{-1}$ . Therefore, the mean activity is equal to

$$\langle \bar{x} \rangle = \frac{g_{ext} \sqrt{K_{ext}}}{1 + g\sqrt{K}} \bar{x}_{ext} \quad (13)$$

This expression is used in the main text (Eq.(4)).

Different neurons have different connections and therefore different activity, and the extent to which the activity varies from neuron to neuron is determined by the spatial variance. We calculate this quantity by taking the sample variance across neurons and averaging over the spatial disorder. We take the scalar product of Eq.(11) with itself, and we use again the fact that the sample mean does not depend on the spatial disorder for large  $N$ , to obtain

---

<sup>3</sup>In the limit of large  $N$ , the real part of the eigenvalues of  $G$  is bounded by  $\lambda$ . Therefore, if  $\lambda \geq 1$ , some eigenvalues of  $(G - I)$  have non-negative real part, and the integral does not converge. This corresponds to an unstable fixed point at  $\mathbf{x} = 0$  and network activity grows in time without bounds.

$$\langle \Delta \bar{x}^2 \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \Delta \bar{x}_i^2 \right\rangle = \left\langle \frac{\bar{x}_{ext}^2}{N} \mathbf{1}^\dagger G_{ext}^\dagger (I - G^\dagger)^{-1} (I - G)^{-1} G_{ext} \mathbf{1} \right\rangle - \langle \bar{x} \rangle^2 \quad (14)$$

We also use the trace operator and its cyclic invariance, to rewrite this expression as

$$\langle \Delta \bar{x}^2 \rangle = \left\langle \frac{\bar{x}_{ext}^2}{N} \text{Tr} \left( (I - G^\dagger)^{-1} (I - G)^{-1} G_{ext} \mathbf{1} \mathbf{1}^\dagger G_{ext}^\dagger \right) \right\rangle - \langle \bar{x} \rangle^2 \quad (15)$$

Again, since  $G$  and  $G_{ext}$  are independent, we can average separately the factors involving the two matrices. A simple calculation shows that  $\langle G_{ext} \mathbf{1} \mathbf{1}^\dagger G_{ext}^\dagger \rangle = g_{ext}^2 K_{ext} \mathbf{1} \mathbf{1}^\dagger + \lambda_{ext}^2 I$ . Furthermore, we show in Appendix 2, Eqs.(48,49) that the following two equalities hold  $\langle \text{Tr}((I - G)^{-1} (I - G^\dagger)^{-1}) \rangle = N(1 - \lambda^2)^{-1}$ , and  $\langle \text{Tr}((I - G^\dagger)^{-1} (I - G)^{-1} \mathbf{1} \mathbf{1}^\dagger) \rangle = N(1 - \lambda^2)^{-1} (1 + g\sqrt{K})^{-2}$ . Using the expression of the mean activity, Eq.(13), the spatial variance is equal to

$$\langle \Delta \bar{x}^2 \rangle = \frac{1}{1 - \lambda^2} \left[ \langle \bar{x} \rangle^2 \lambda^2 + \bar{x}_{ext}^2 \lambda_{ext}^2 \right] \quad (16)$$

This expression is used in the main text (Eq.(5)).

After looking at the spatial variability, we study temporal variability and correlated fluctuations by calculating the covariance matrix. We take the scalar product of Eq.(10) with itself and we perform the temporal average, using the fact that the external stimulus is uncorrelated in space and time. This corresponds to the covariance matrix of a Ornstein-Uhlenbeck process [16], and is equal to

$$Q = \overline{\Delta \mathbf{x} \Delta \mathbf{x}^\dagger} = \overline{\Delta x_{ext}^2} \int_0^{+\infty} dt e^{(G-I)t} G_{ext} G_{ext}^\dagger e^{(G^\dagger-I)t} \quad (17)$$

Note that the covariance matrix satisfies the equation  $(G - I)Q + Q(G^\dagger - I) = G_{ext} G_{ext}^\dagger$ , but this cannot be used for averaging  $Q$  since  $G$  and  $Q$  are dependent and they do not commute. Also  $G$  and  $G_{ext} G_{ext}^\dagger$  do not commute. The on-diagonal elements of the covariance matrix are the temporal variances of different neurons. To determine the average temporal variance, we take the sample mean across neurons and we average over the spatial disorder, obtaining

$$\langle \overline{\Delta x^2} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \overline{\Delta x_i^2} \right\rangle = \frac{\overline{\Delta x_{ext}^2}}{N} \int_0^{+\infty} dt \langle \text{Tr} \left( e^{(G^\dagger-I)t} e^{(G-I)t} G_{ext} G_{ext}^\dagger \right) \rangle \quad (18)$$

where we applied the trace operator and used its cyclic invariance. A simple calculation gives  $\langle G_{ext} G_{ext}^\dagger \rangle = k_{ext} g_{ext}^2 \mathbf{1} \mathbf{1}^\dagger + \lambda_{ext}^2 I$ . Furthermore, using Eqs(50,51) in Appendix 2, we obtain

$$\langle \overline{\Delta x^2} \rangle = \frac{\overline{\Delta x_{ext}^2}}{2\sqrt{1 - \lambda^2}} \left\{ \frac{k_{ext} g_{ext}^2}{1 + g\sqrt{K}} \left[ \frac{1 + \sqrt{1 - \lambda^2} (1 + g\sqrt{K})}{1 + \sqrt{1 - \lambda^2} + g\sqrt{K}} \right] + \lambda_{ext}^2 \right\} \quad (19)$$



In the main text, we substitute the term in square brackets with its limit for large  $g\sqrt{K}$ , which is equal to  $\sqrt{1-\lambda^2}$ . The result is Eq.(6).

Next, we calculate the average covariance, by looking at the off-diagonal elements of the matrix in Eq.(17). The off-diagonal elements are the pairwise covariances, and the sample mean across neuron pairs can be averaged over the spatial disorder to obtain the average covariance

$$\begin{aligned}\langle \overline{\Delta x \Delta x} \rangle &= \left\langle \frac{1}{N(N-1)} \sum_{i \neq j}^{1,N} \overline{\Delta x_i \Delta x_j} \right\rangle = \\ &= \frac{\overline{\Delta x_{ext}^2}}{N(N-1)} \int_0^{+\infty} dt \left\langle \text{Tr} \left( e^{(G^\dagger - I)t} (\mathbf{1}\mathbf{1}^\dagger - I) e^{(G - I)t} G_{ext} G_{ext}^\dagger \right) \right\rangle\end{aligned}$$

Using Eq.(52) in Appendix 2, we obtain

$$\langle \overline{\Delta x \Delta x} \rangle = \frac{\overline{\Delta x_{ext}^2}}{2} \frac{k_{ext} g_{ext}^2}{1 + g\sqrt{K}} \quad (20)$$

This expression is used in the main text (Eq.(7)).

The mean correlation is obtained by dividing the covariance, Eq.(20), by the variance, Eq.(19), i.e. (we assume that variance and covariance are independent):

$$\langle R \rangle = \frac{\langle \overline{\Delta x \Delta x} \rangle}{\langle \overline{\Delta x^2} \rangle} = \frac{\sqrt{1-\lambda^2}}{\frac{1+\sqrt{1-\lambda^2}(1+g\sqrt{K})}{1+\sqrt{1-\lambda^2}+g\sqrt{K}} + \frac{\lambda_{ext}^2(1+g\sqrt{K})}{k_{ext}g_{ext}^2}} \quad (21)$$

This expression is positive and never exceeds one. In the main text, we substitute the first term in the denominator with its limit for large  $g\sqrt{K}$ , which is equal to  $\sqrt{1-\lambda^2}$ . The result is Eq.(8). Note that, due to this simplification, the equation in the main text may be inaccurate in some cases. For example, if the external heterogeneity is removed,  $\lambda_{ext} = 0$ , then Eq.(8) gives  $\langle R \rangle = 1$ , while Eq.(21) gives

$$\langle R \rangle = 1 - \frac{\lambda^2}{1 + \sqrt{1-\lambda^2}(1 + g\sqrt{K})} \quad (22)$$

This result is more accurate since, if the local heterogeneity is not removed,  $\lambda \neq 0$ , then the network is not precisely homogeneous and the correlation is not expected to be exactly equal to one (however, correlation is one for large  $K$  even in Eq.(22)). The two expressions also differ for  $g = 0$ .

## Appendix 2: Traces of random matrix products

In this section we introduce the diagrammatic notation to calculate the quenched averages of random matrix products (see e.g. [20]). Theoretical results are obtained for the Gaussian distribution, although numerical simulations suggest that



they generalize to other distributions with the same mean and variance (e.g. Bernouilli). We will consider the case in which the mean of the matrix element is  $\sim g/N$  and then recover the scaling studied in the main text by analytical continuation and the substitution  $g \rightarrow g\sqrt{K}$ . We conclude the section by studying the case of non-homogeneous mean (e.g. interconnected excitatory and inhibitory neurons).

We start with the problem of calculating the quenched average of the trace of a power of the random matrix  $R$  in the limit of large  $N$  (where the size of the matrix is  $N \times N$ ). The matrix  $R$  is characterized by independent and normally distributed elements, each element having zero mean and variance  $N^{-1}$ , namely

$$\langle R_{ij} \rangle = 0 \quad \langle R_{ij}^2 \rangle = \frac{1}{N} \quad (23)$$

We start by calculating the second order, i.e. the average trace of the square of  $R$ . For convenience of notation, we omit the sum over the indices (in this case the sum over the indices  $a, b, c, d$ )

$$\langle \text{Tr}(R^2) \rangle = \delta_{ad} \delta_{bc} \langle R_{ab} R_{cd} \rangle = N^{-1} \delta_{ad} \delta_{bc} \delta_{ac} \delta_{bd} \quad (24)$$

The diagram corresponding to this expression is shown in Fig.5a. The diagram is obtained by drawing one node for each one of the four indices  $a, b, c, d$ , and by drawing an edge for each delta function in the expression, where the two nodes connected by the edge correspond to the two indices of the delta function. Horizontal edges are due to the operations of trace (base edge) and matrix multiplication (middle edge), while arc-shaped edges are due to averaging. The multiple edges determine different paths, and each pair of nodes connected by a path (even if not linked by an edge), corresponds to a pair of indices that must be equal, since they are connected by a sequence of delta functions. Therefore, for each closed loop in the diagram there is one redundant delta function, which can be eliminated without performing the sum over the corresponding indices. This implies that each closed loop contributes with a factor  $N$ , due to a free sum over  $N$  elements. Since the diagram for the second order has one loop, we have  $\langle \text{Tr}(R^2) \rangle = N^{-1} N = 1$ .

Note that all terms of odd order are zero, because  $\langle R_{ij}^k \rangle = 0$  for odd  $k$ . The next order is therefore the fourth order, which is equal to (again we omit the sum over all indices)

$$\langle \text{Tr}(R^4) \rangle = \delta_{ah} \delta_{bc} \delta_{de} \delta_{fg} \langle R_{ab} R_{cd} R_{ef} R_{gh} \rangle = \quad (25)$$

$$= N^{-2} \delta_{ah} \delta_{bc} \delta_{de} \delta_{fg} [\delta_{ac} \delta_{bd} \delta_{eg} \delta_{fh} + \delta_{ae} \delta_{bf} \delta_{cg} \delta_{dh} + \delta_{ag} \delta_{bh} \delta_{ce} \delta_{df}] \quad (26)$$

The fourth order has three diagrams, one for each term in the sum, shown in Fig.5b. The middle diagram has two closed loops, while the other two have only one loop. Therefore the other two terms can be neglected, the middle term

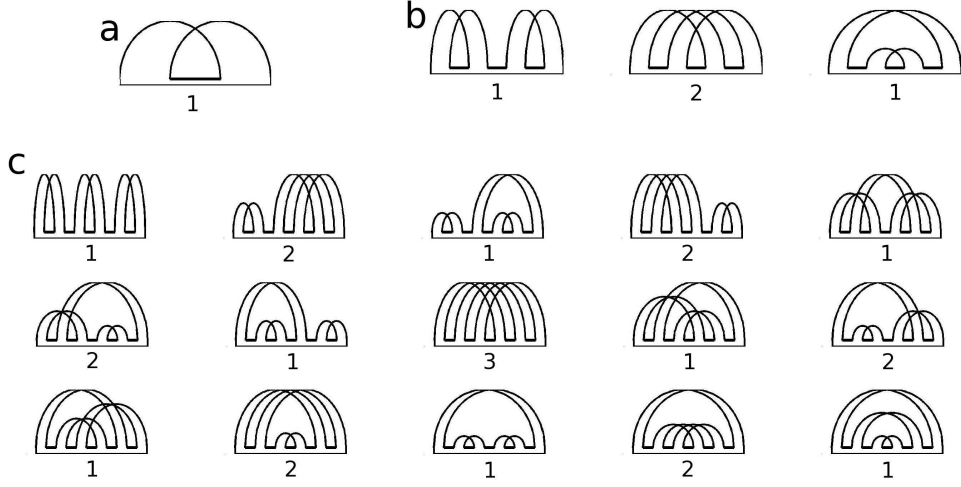


Figure 5: Diagrams of the traces of random matrix powers, described by Eq.(27). Below each diagram the number of its closed loops is indicated. a) Second order. b) Fourth order. c) Sixth order.

contributes with a factor  $N^2$  and the fourth order gives  $\langle \text{Tr}(R^4) \rangle = 1$ . The contribution of the fourth order moment ( $\langle R_{ij}^4 \rangle$ ) can be neglected.

The fifteen diagrams of the sixth order are shown in Fig.5c. Again, we neglect moments higher than the second, and we note that only one diagram contributes with three loops, therefore  $\langle \text{Tr}(R^6) \rangle = 1$ . By iterating this procedure, we find that order  $2k$  has  $(2k - 1)!!$  diagrams of which only one has  $k$  loops, therefore

$$\langle \text{Tr}(R^{2k}) \rangle = 1 \quad (27)$$

for all values of  $k$ .

Note that the elements of the matrix  $R$  have zero mean, while the matrices considered in the main text ( $G$  and  $G_{ext}$ ) have non-zero mean. As explained below, in order to calculate the average trace of matrix powers with non-zero mean, we need to compute averages where  $R$  is interleaved by the matrix of ones. We denote by  $\mathbf{1}$  the column vector of  $N$  components all equal to one, by  $\mathbf{1}^\dagger$  the row vector, and by  $\mathbf{1}\mathbf{1}^\dagger$  the  $N \times N$  matrix with all elements equal to one (we denote by  $\dagger$  the transpose operation). We consider the two second order terms

$$\langle \text{Tr}(R\mathbf{1}\mathbf{1}^\dagger R) \rangle = \delta_{ad} \langle R_{ab}R_{cd} \rangle = N^{-1} \delta_{ad} \delta_{ac} \delta_{bd} = 1 \quad (28)$$

$$\langle \text{Tr}(R^2\mathbf{1}\mathbf{1}^\dagger) \rangle = \delta_{bc} \langle R_{ab}R_{cd} \rangle = N^{-1} \delta_{bc} \delta_{ac} \delta_{bd} = 1 \quad (29)$$

It is not surprising that these two expressions are equal, since the trace is cyclic invariant. The only difference of these expressions with Eq.(24) is the absence of a factor  $\delta_{bc}$  in the former expression, and  $\delta_{ad}$  in the latter. This corresponds to cutting, respectively, the middle and the base horizontal edges in the diagram of

Fig.5a. In general, inserting a matrix of ones at a given point of the sequence of  $R$  products is equivalent to cutting the horizontal edge at that point in the corresponding diagram. If the edge belongs to a closed loop, the cut has the only effect of removing a redundant delta function, and there is no change in the contribution of that diagram to the sum; Conversely, if the edge belongs to an open path, the cut determines an additional  $N$  factor, because the delta function removed was not redundant. Since all diagrams have at least one closed loop, inserting a single matrix of ones has no effect at all orders. Therefore,

$$\langle \text{Tr} (R^{2k-k'} \mathbf{1}\mathbf{1}^\dagger R^{k'}) \rangle = 1 \quad (30)$$

for all  $k' = 0, \dots, 2k$ . Unless more loops are available to cut, inserting more matrices of ones may cut open paths, therefore the trace may be multiplied by  $N$ . An additional  $N$  factor is obtained also by multiplying the matrix of ones with itself, which occurs whenever additional matrices are inserted at same point in the sequence (we have that  $\mathbf{1}^\dagger \mathbf{1} = N$  and  $(\mathbf{1}\mathbf{1}^\dagger)^k = N^{k-1} \mathbf{1}\mathbf{1}^\dagger$  if  $k > 0$ ).

Using the above results, we can calculate the average trace of random matrix powers with non-zero mean and arbitrary variance (provided that the variance is of order  $N^{-1}$ ). We consider the matrix  $G$  equal to

$$G = \frac{g}{N} \mathbf{1}\mathbf{1}^\dagger + \lambda R \quad (31)$$

Note that the mean of this matrix has a different scaling with respect to that considered in the main text, but we will recover the latter by the substitution  $g \rightarrow -\sqrt{K}g$ . A power of  $G$  is calculated by multiplying  $G$  to itself, and this determines an ordered product of powers of the matrices  $R$  and  $\mathbf{1}\mathbf{1}^\dagger$ . Note that these two matrices do not commute, therefore the binomial theorem cannot be applied. We consider the average trace

$$\langle \text{Tr} (G^k) \rangle = \sum_{k'=0}^k N^{-k'} g^{k'} \lambda^{k-k'} \sum \binom{k}{k'} \langle \text{Tr}(\dots) \rangle \quad (32)$$

where the trace in the right hand side is applied to an ordered product of  $k'$  matrices  $\mathbf{1}\mathbf{1}^\dagger$  and  $k - k'$  matrices  $R$ , and the sum runs over all the  $\binom{k}{k'}$  ordered products for a given  $k$  and  $k'$ . Using the above results, we find that the contribution of any of those traces is zero for  $k - k'$  odd, is equal to one for  $k' = 0$  (provided that  $k$  is even), is equal to  $N^k$  for  $k' = k$  and is at most of order  $N^{k'-1}$  for  $k' = 1, \dots, k - 1$ . Therefore, the leading order terms are  $k' = k$  (for any value of  $k$ ) and  $k' = 0$  (for  $k$  even), and all other terms can be neglected. If the matrix  $G^k$  is further multiplied by a matrix of ones, the term  $k' = 0$  can also be neglected, and we find that

$$\langle \text{Tr} (G^k \mathbf{1}\mathbf{1}^\dagger) \rangle = N g^k = \text{Tr} (\langle G \rangle^k \mathbf{1}\mathbf{1}^\dagger) \quad (33)$$

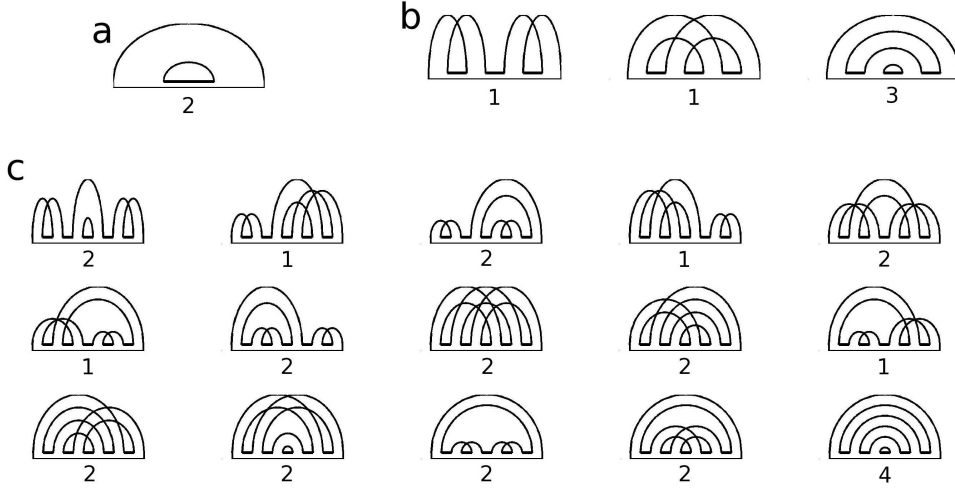


Figure 6: Diagrams of the traces of random matrix powers multiplied by its transpose, described by Eq.(40). Below each diagram the number of its closed loops is indicated. a) Second order. b) Fourth order. c) Sixth order.

for all values of  $k$ . Note that if the mean of  $G$  has a higher order in  $N$ , the result still holds. This expression is particularly useful to compute the average of bracket expressions. Because  $\text{Tr}(A\mathbf{x}\mathbf{y}^\dagger) = \mathbf{y}^\dagger A\mathbf{x}$  for any matrix  $A$  and vectors  $\mathbf{x}$ ,  $\mathbf{y}$ , the expression can be rewritten as

$$\langle \mathbf{1}^\dagger G^k \mathbf{1} \rangle = \mathbf{1}^\dagger \langle G \rangle^k \mathbf{1} \quad (34)$$

for all values of  $k$ . Since any infinitely differentiable function  $f$  can be expanded in Taylor series, the above result implies that

$$\langle \mathbf{1}^\dagger f(G) \mathbf{1} \rangle = \mathbf{1}^\dagger f(\langle G \rangle) \mathbf{1} \quad (35)$$

Therefore, the following expression can be calculated and used to compute the mean activity in the main text,

$$\langle \mathbf{1}^\dagger (I - G)^{-1} \mathbf{1} \rangle = \frac{N}{1 - g} \quad (36)$$

Note that the substitution  $g \rightarrow -g\sqrt{K}$  must be applied to recover the scaling studied in the main text.

Next, we calculate the diagrammatic expansion for products of a random matrix with its transpose. Again, all odd orders vanish and we neglect moments higher than the second at all orders. The second order term is

$$\langle \text{Tr}(RR^\dagger) \rangle = \delta_{ad}\delta_{bc} \langle R_{ab}R_{dc} \rangle = N^{-1} \delta_{ad}\delta_{bc}\delta_{ad}\delta_{bc} \quad (37)$$

The corresponding diagram has two loops and is shown in Fig.6a, therefore the loops contribute with a factor  $N^2$  and the second order is  $\langle \text{Tr}(RR^\dagger) \rangle = N$ . The fourth order is equal to

$$\langle \text{Tr}(R^2 R^{2\dagger}) \rangle = \delta_{ah} \delta_{bc} \delta_{de} \delta_{fg} \langle R_{ab} R_{cd} R_{fe} R_{hg} \rangle = \quad (38)$$

$$= N^{-2} \delta_{ah} \delta_{bc} \delta_{de} \delta_{fg} [\delta_{ac} \delta_{bd} \delta_{fh} \delta_{eg} + \delta_{af} \delta_{be} \delta_{ch} \delta_{dg} + \delta_{ah} \delta_{bg} \delta_{cf} \delta_{de}] \quad (39)$$

The three diagrams are shown in Fig.6b. The first two diagrams have one loop, while the third has three. Therefore, that diagram contributes with a factor  $N^3$  and the fourth order is equal to  $\langle \text{Tr}(R^2 R^{2\dagger}) \rangle = N$ . The diagrams for the sixth order are shown in Fig.6c: only one diagram has four loops, and no diagram has three, therefore  $\langle \text{Tr}(R^3 R^{3\dagger}) \rangle = N$ . Iterating the procedure, we find that order  $2k$  has  $(2k-1)!!$  diagrams of which only one has  $k+1$  loops, therefore

$$\langle \text{Tr}(R^k R^{k\dagger}) \rangle = N \quad (40)$$

for all values of  $k$ . Other combinations of powers of  $R$  and its transpose give a smaller contribution, i.e.  $\langle \text{Tr}(R^{2k-k'} R^{k'\dagger}) \rangle = o(1)$  for  $k' \neq k$ .

Inserting matrices of ones in this case has a similar effect as in the case above, Eq.(30), each matrix cuts the horizontal edge corresponding to where the matrix is placed. Again, since each diagram has at least one loop, the insertion of a single matrix of ones (and the consequent edge removal) has no effect on the trace at all orders. Therefore,

$$\langle \text{Tr}(R^{k-k'} \mathbf{1} \mathbf{1}^\dagger R^{k'} R^{k\dagger}) \rangle = \langle \text{Tr}(R^k R^{k'\dagger} \mathbf{1} \mathbf{1}^\dagger R^{k-k'\dagger}) \rangle = N \quad (41)$$

An insertion in a term with unequal powers of  $R$  and  $R^\dagger$  remains of order one. Adding more matrices increases the trace by an order  $N$  for each matrix, provided that no further loops are cutted.

Using the above expressions, we can compute the average of products of powers of the matrix  $G$  and its transpose, namely

$$\langle \text{Tr}(G^k G^{l\dagger}) \rangle = \sum_{k'=0}^k \sum_{l'=0}^l N^{-k'-l'} g^{k'+l'} \lambda^{k+l-k'-l'} \sum_{k'=0}^k \sum_{l'=0}^l \langle \text{Tr}(\dots) \rangle \quad (42)$$

where the trace in the right hand side is applied to an ordered product of  $k'+l'$  matrices  $\mathbf{1} \mathbf{1}^\dagger$ ,  $k-k'$  matrices  $R$  and  $l-l'$  matrices  $R^\dagger$ . If  $k'=k$  and  $l'=l$ , the trace is equal to  $N^{k+l}$ , and the term is of order one. If  $k'=0$  and  $l'=0$ , the trace contributes with an order  $N$ , provided that  $k=l$ . If  $k-k'=l-l'$ , the traces contribute at most with an order  $N^{k'+l'}$ , and the term is of order one, while if  $k-k' \neq l-l'$  the term is of smaller order. Therefore the leading order is  $N$ , and we have

$$\langle \text{Tr} (G^k G^{l\dagger}) \rangle = N \delta_{kl} \lambda^{k+l} \quad (43)$$

In the case in which matrices of ones are inserted, the term  $k' = 0, l' = 0$  is no longer leading, and many other terms have to be considered. Those are the terms for  $k - k' = l - l'$ , and for which additional inserted matrices cuts the same loop. Since the leading diagrams at all orders have one two-nodes loop in the middle and one at the boundaries, if a matrix of ones is inserted in the middle or at the boundaries, additional matrices must continue to be inserted at the same place in order to cut the same loop. We eliminate one sum and we use the index  $m = k - k' = l - l'$  in place of  $k'$  and  $l'$ ; We obtain

$$\langle \text{Tr} (G^k G^{l\dagger} \mathbf{1} \mathbf{1}^\dagger) \rangle = \sum_{m=0}^{\min(k,l)} N^{2m-k-l} g^{l+k-2m} \lambda^{2m} \langle \text{Tr} ((\mathbf{1} \mathbf{1}^\dagger)^{k-m} R^m R^{m\dagger} (\mathbf{1} \mathbf{1}^\dagger)^{l-m+1}) \rangle \quad (44)$$

$$\langle \text{Tr} (G^k \mathbf{1} \mathbf{1}^\dagger G^{l\dagger}) \rangle = \sum_{m=0}^{\min(k,l)} N^{2m-k-l} g^{l+k-2m} \lambda^{2m} \langle \text{Tr} (R^m (\mathbf{1} \mathbf{1}^\dagger)^{k+l-2m+1} R^{m\dagger}) \rangle \quad (45)$$

Both expressions are equal to

$$\langle \text{Tr} (G^k G^{l\dagger} \mathbf{1} \mathbf{1}^\dagger) \rangle = \langle \text{Tr} (G^k \mathbf{1} \mathbf{1}^\dagger G^{l\dagger}) \rangle = N \sum_{m=0}^{\min(k,l)} g^{l+k-2m} \lambda^{2m} \quad (46)$$

Furthermore, we calculate the average trace with two inserted matrices. In that case, the leading term is for  $k = k'$  and  $l = l'$  (or  $m = 0$ ), and we obtain

$$\langle \text{Tr} (G^k \mathbf{1} \mathbf{1}^\dagger G^{l\dagger} \mathbf{1} \mathbf{1}^\dagger) \rangle = N^2 g^{k+l} = \text{Tr} (\langle G \rangle^k \mathbf{1} \mathbf{1}^\dagger \langle G \rangle^{l\dagger} \mathbf{1} \mathbf{1}^\dagger) \quad (47)$$

Using the expressions above and the Taylor series expansion of infinitely differentiable functions, we calculate the following traces that are used in the main text to compute the variance and covariance of the activity

$$\langle \text{Tr} ((I - G)^{-1} (I - G^\dagger)^{-1}) \rangle = \frac{N}{1 - \lambda^2} \quad (48)$$

$$\langle \text{Tr} ((I - G)^{-1} \mathbf{1} \mathbf{1}^\dagger (I - G^\dagger)^{-1}) \rangle = \frac{N}{(1 - \lambda^2)(1 - g)^2} \quad (49)$$

$$\int_0^\infty dt e^{-2t} \langle \text{Tr} (e^{Gt} e^{G^\dagger t}) \rangle = \frac{N}{2\sqrt{1 - \lambda^2}} \quad (50)$$

$$\int_0^\infty dt e^{-2t} \langle \text{Tr} (e^{Gt} \mathbf{1} \mathbf{1}^\dagger e^{G^\dagger t}) \rangle = \frac{N}{2\sqrt{1 - \lambda^2}(1 - g)} \left[ \frac{1 + \sqrt{1 - \lambda^2}(1 - g)}{1 + \sqrt{1 - \lambda^2} - g} \right] \quad (51)$$

$$\int_0^\infty dt e^{-2t} \langle \text{Tr} (e^{Gt} \mathbf{11}^\dagger e^{G^\dagger t} \mathbf{11}^\dagger) \rangle = \frac{N^2}{2(1-g)} \quad (52)$$

Note that the substitution  $g \rightarrow -g\sqrt{K}$  must be applied to recover the scaling studied in the main text. If  $K$  is proportional to  $N$ , this substitution may change the order of magnitude of various terms in the summation considered above, possibly modifying the leading terms in each sum. Note that all series converge only for  $|g| < 1$ , but their sum can be evaluated at  $g \rightarrow -g\sqrt{K}$  by analytical continuation. Then, approximating the sums by the leading terms described above is accurate under the assumption that all series involving lower order terms converge to bounded functions of  $g$ .

We conclude this section by noting that the above calculations can be performed also in the case of a non-homogeneous mean. We have assumed that the mean matrix is homogeneous, i.e. is proportional to the matrix of ones  $\langle G \rangle \propto \mathbf{11}^\dagger$ . However, the same approach could be taken to analyze the more general case in which the mean is a rank-1 matrix, i.e.  $\langle G \rangle = \mathbf{ab}^\dagger$ , where  $\mathbf{a}$  and  $\mathbf{b}$  are two arbitrary vectors. For example, if the mean connection strength only depends on the pre-synaptic neuron,  $\langle G \rangle = \mathbf{1b}^\dagger$ , then interconnected excitatory and inhibitory neurons can be considered, where the vector  $\mathbf{b}$  is positive for excitatory neurons and negative for the inhibitory ones. In that case, all results above still hold, with the simple substitution  $g \rightarrow N^{-1}\mathbf{b}^\dagger \mathbf{1}$ .

## References

- [1] Abbott LF, Dayan P (1999) The Effect of Correlated Variability on the Accuracy of a Population Code. *Neural Computation* 11, 911-101.
- [2] Averbeck BB, Lee D (2003) Neural Noise and Movement-Related Codes in the Macaque Supplementary Motor Area. *J. Neurosci.* 23:7630-7641.
- [3] Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation, *Nat. Rev. Neurosci.* 7:358-366.
- [4] Bair W, Zohary E, Newsome WT (2001) Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior, *J. Neurosci.* 21:1676-1697.
- [5] Bernacchia A, Amit DJ (2007) Impact of spatiotemporally correlated images on the structure of memory. *Proc. Natl. Acad. Sci. USA* 104:3544-3549.
- [6] Bernacchia A, Seo H, Lee D, Wang XJ (2011) A reservoir of time constants for memory traces in cortical neurons. *Nature Neurosci.* 14:366-372.

- [7] Braitenberg V, Schuz A. *Anatomy of the Cortex: Statistics and Geometry*. Springer-Verlag, 1991.
- [8] Cafaro Jon, Rieke F (2010) Noise correlations improve response fidelity and stimulus encoding. *Nature* 468:974-967.
- [9] Cohen MR, Newsome WT (2008) Context-Dependent Changes in Functional Circuitry in Visual Area MT, *Neuron* 60:172-163.
- [10] Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations, *Nature Neurosci.* 12:1594-1600.
- [11] Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nature Neurosci.* 14:811-819.
- [12] Constantinidis C, Goldman-Rakic PS (2002) Correlated Discharges Among Putative Pyramidal Neurons and Interneurons in the Primate Prefrontal Cortex, *J. Neurophysiol* 88: 3487-3497.
- [13] de la Rocha J, Doiron B, Shea-Brown E, Josic K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. *Nature* 448:802-806.
- [14] Destexhe A, Rudolph M, Par D (2003) The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci.* 4:739-751.
- [15] Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated Neuronal Firing in Cortical Microcircuits, *Science* 327:584-587.
- [16] Gardiner CW, *Handbook of stochastic methods: for Physics, Chemistry and Natural Sciences*. Springer; 2nd edition (January 1985).
- [17] Ginzburg I, Sompolinski H (1994) Theory of correlations in stochastic neural networks. *Phys Rev. E* 50:3171-3191.
- [18] Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neurosci.* 14:239-245.
- [19] Gutniski DA, Dragoi V (2008) Adaptive coding of visual information in neural populations, *Nature* 452:220-224.
- [20] Gudowska-Nowak E, Janik RA, Jurkiewicz J, Nowak MA (2003) Infinite products of large random matrices and matrix-valued diffusion. *Nuclear Physics B* 670:479507.
- [21] Hertz J, Cross-Correlations in High-Conductance States of a Model Cortical Network, *Neural Computation* 22, 427-447.



- [22] Huang X, Lisberger SG (2009) Noise Correlations in Cortical Area MT and Their Potential Impact on Trial-by-Trial Variation in the Direction and Speed of Smooth-Pursuit Eye Movements, *J. Neurophysiol.* 101: 3012-3030.
- [23] Kohn A, Smith MA (2005) Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. *J. Neurosci.* 25:3661-3673.
- [24] Komiyama T, Sato TR, OConnor DH, Zhang Y-X, Huber D, Hooks BM, Gabitto M, Svoboda K (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* 464:1182-1186.
- [25] Lampl I, Reichova I, Ferster D (1999) Synchronous Membrane Potential Fluctuations in Neurons of the Cat Visual Cortex. *Neuron* 22:361-374.
- [26] Lee D, Port NL, Kruse W, Georgopoulos AP (1998) Variability and Correlated Noise in the Discharge of Neurons in Motor and Parietal Areas of the Primate Cortex. *J. Neurosci.* 18:1161-1170.
- [27] Maynard EM, Hatsopoulos NG, Ojakangas CL, Acuna BD, Sanes JN, Normann RA, Donoghue JP (1999) Neuronal Interactions Improve Cortical Population Coding of Movement Direction. *J. Neurosci.*, 19:8083-8093.
- [28] Mitchell JF, Sundberg KA, Reynolds JH (2010) Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron* 63, 879-888.
- [29] Nadal JP, Parga N (1994) Non-linear neurons in the low-noise limit: a factorial code maximizes information transfer. *Network: Comp. Neu. Syst.* 5:565-581.
- [30] Okun M, Lampl I (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neurosci.* 11:535-537.
- [31] Panzeri S, Schultz SR, Treves A, Rolls ET (1999) Correlations and the encoding of information in the nervous system. *Proc. R. Soc. Lond. B* 266:1001-1012.
- [32] Poulet JFA, Petersen CC (2008) Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* 454:881-885.
- [33] Raz A, Vaadia E, Bergman H (2000) Firing Patterns and Correlations of Spontaneous Discharge of Pallidal Neurons in the Normal and the Tremulous 1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine Vervet Model of Parkinsonism. *J. Neurosci.* 20:8559-8571.

- [34] Reich DS, Mechler F, Victor JD (2001) Independent and Redundant Information in Nearby Cortical Neurons. *Science* 294:2566-2568.
- [35] Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD (2010) The Asynchronous State in Cortical Circuits. *Science* 327:587-590.
- [36] Romo R, Hernandez A, Zainos A, Salinas E (2003) Correlated Neuronal Discharges that Increase Coding Efficiency during Perceptual Discrimination. *Neuron* 38:649-657.
- [37] Salinas E, Sejnowski TJ (2000) Impact of Correlated Synaptic Input on Output Firing Rate and Variability in Simple Neuronal Models. *J. Neurosci.* 20:6193-6209.
- [38] Shadlen MN, Newsome WT, The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *J. Neurosci.* 18:3870-3896.
- [39] Smith MA, Kohn A (2008) Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex, *J. Neurosci.* 28:12591-12603.
- [40] Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys. Rev. E* 64:051904.
- [41] van Vreeswijk CA, Sompolinski H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274:1724-1726.
- [42] Wilke SD, Eurich CW (2002) On the functional role of noise correlations in the nervous system. *Neurocomputing* 44-46, 1023-1028.
- [43] Wilson CJ, Beverlin II B, Netoff T (2011) Chaotic desynchronization as the therapeutic mechanism of deep brain stimulation. *Front. Syst. Neurosci.* 5:50.
- [44] Zohary E, Shadlen MN, Newsome W (1994) Correlated neuronal discharge rate and its implications for psychophysical performance, *Nature* 370:140-143